

InAR:Inverse Augmented Reality

Hao Hu
Department of Electronics
Shan-Dong University
Jinan, China
huhao678@126.com

Hainan Cui
Institute of Automation
Chinese Academy of Sciences
Beijing, China
hncui@nlpr.ia.ac.cn

Abstract

Augmented reality is the art to seamlessly fuse virtual objects into real ones. In this short note, we address the opposite problem, the inverse augmented reality, that is, given a perfectly augmented reality scene where human is unable to distinguish real objects from virtual ones, how the machine could help do the job. We show by structure from motion (SFM), a simple 3D reconstruction technique from images in computer vision, the real and virtual objects can be easily separated in the reconstructed 3D scene.

1. Introduction

Augmented reality (AR) is meant to seamlessly fuse interested virtual objects into the real scene, and one of its ultimate goals is to make the fusion imperceptible even by an experienced human being, and great efforts and progresses have been made on it in AR community in recent years [4].

In this work, we address the opposite problem: suppose one emerges in a perfectly fused AR environment where he is unable to distinguish which one is virtual, which one is real around him by his own visual system, how could he resort to a machine, for example, his mobile phone or a common digital camera, to aid him to ascertain virtual from real ones? To our knowledge, this inverse problem seems unaddressed in the literature. Here we show by a SFM-based simple 3D reconstruction technique from images captured by his mobile phone camera, the job can be easily done.

2. Inverse Augmented Reality

Shape-from-Motion (SFM) is a widely used technique in 3D scene reconstruction from images in computer vision field. By establishing point correspondences across several images and by some assumptions on the camera's intrinsic parameters, currently a dense metric scene reconstruction can be automatically obtained, for example, by Snavely's Bundler [3] for sparse reconstruction, followed

by Furukawa's PMVS [1] for dense or quasi-dense reconstruction.

In the following, we report a case to illustrate the basic principle and effects. Shui-Dong-Gou Archeological Museum is located in the Ning-Xia-Hui Autonomous Region, North West China, it is an archeological site of human life and activities dating back to about 40,000 years. An example image of this museum is showed in Fig. 1(a). Currently its demo hall is in such an exploit of AR technology that ordinary visitors can hardly tell which part of scene is virtual, which part is real.

As shown in Fig. 1(a), we can hardly distinguish virtual parts from real scene, at least for the majority of the lay visitors. By taking about 20 images with a Canon 5D Mark III camera, and by Bundler + PMVS2 combination, the AR scene is 3D reconstructed as shown in Fig. 1(b) and Fig. 1(c). Note that in Fig. 1(b) and Fig. 1(c), small colored cones are calibrated camera poses of the used images for our 3D reconstruction, and the Fig. 1(c) is the top-view of dense reconstruction. From Fig. 1(c), we can see that the sky marked by ① in Fig. 1(a) is in fact a vaulted backdrop, and the greensward marked by ② in Fig. 1(a) is also in this backdrop. Stones marked by ③④ are real ones. In sum, virtual parts can be clearly separated from the real scene from the reconstructed 3D scene.

3. Conclusion

Before ending this short note, we would make the following two points: (1) In this work, we only discuss a static AR scene, the principle and techniques can be extended straightforwardly to dynamic AR scene using a synchronized stereo system. For example, nowadays many mobile phones are equipped with a micro-array camera, which could capture several images at the same time. From such synchronized images, the same 3D reconstruction technique for the static scene can be used for dynamic scenes. (2) Human 3D visual perception is not necessarily truly 3D in the physical sense although the machine and human use similar mechanism for disparity computation [2].

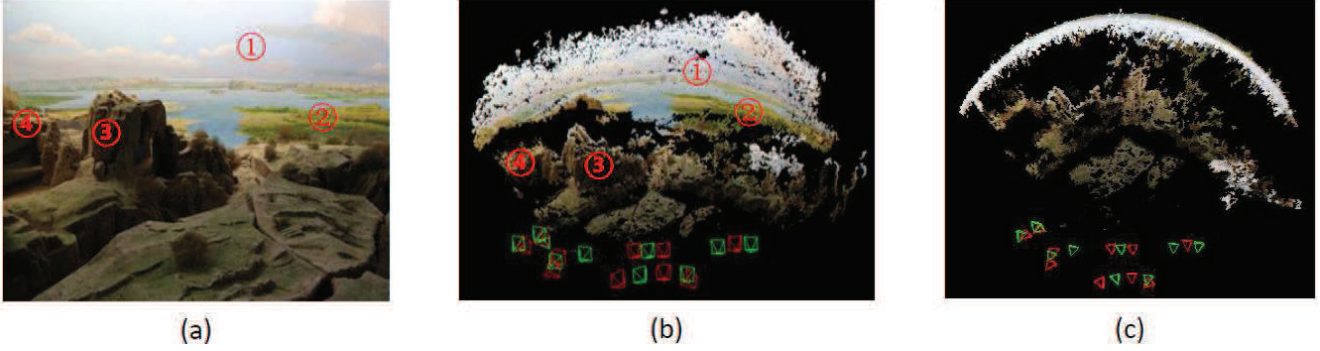


Figure 1. (a): an image of Shui-Dong-Gou Archeological Museum; (b): the dense reconstruction of Shui-Dong-Gou Archeological Museum produced by Bundler+PMVS2; (c): the top-view of the dense reconstruction scene in (b). Note that labels ①②③④ in (a) mark some areas in the scene, and their corresponding reconstruction results are respectively showed in (b).

References

- [1] Y. Furukawa and J. Ponce. Accurate, dense, and robust multiview stereopsis. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 32(8):1362–1376, 2010. [1](#)
- [2] F. R. Ohzawa I, DeAngelis G. Stereoscopic depth discrimination in the visual cortex: neurons ideally suited as disparity detectors. *Science*, 4972(249):1037–1041, 1990. [1](#)
- [3] N. Snavely, S. M. Seitz, and R. Szeliski. Modeling the world from internet photo collections. *International Journal of Computer Vision*, 80(2):189–210, 2008. [1](#)
- [4] D. Van Krevele and R. Poelman. A survey of augmented reality technologies, applications and limitations. *International Journal of Virtual Reality*, 2(9):1–20, 2010. [1](#)